United States General Accounting Office

GAO

Testimony

Before the Subcommittee on Commerce,
Consumer, and Monetary Affairs
Committee on Government Operations
House of Representatives

For Release
on Delivery
Expected at
9:30 a.m. EDT
Wednesday
April 29, 1992

TAX SYSTEMS MODERNIZATION

Input Processing Strategy is Risky and Lacks a Sound Analytical Basis

Statement of Howard G. Rhile, Director
General Government Information Systems
Information Management and Technology Division

Mr. Chairman and Members of the Subcommittee:

We are pleased to be here today to discuss our work for the Subcommittee relating to the Internal Revenue Service's (IRS) integrated input processing initiative. This initiative is part of IRS' multi-year, multibillion dollar Tax Systems Modernization program and is intended to allow the agency to drastically reduce the manual processes associated with handling paper income tax returns, tax payments, information returns, and other correspondence. This reduction is to be accomplished, in part, by using imaging and optical character recognition (OCR) to electronically capture and optically read data on paper documents. Any subsequent work with the data would be done entirely electronically. As you requested, our work addressed four questions:

(1) Has IRS adequately assessed the cost and benefit tradeoffs of receiving and imaging paper returns and documents compared to receiving and processing input in electronic form?

(2) If IRS were to aggressively market various forms of electronic input, thereby reducing the percentage of input requiring imaging, is there a reasonable likelihood that imaging would continue to be cost effective for IRS?

(3) Is IRS' current input processing strategy technically feasible and realistic given the state of imaging/character recognition technology and the complexities of implementing this technology on IRS' scale?

(4) How are the two current tax return imaging projects, the Document Processing System (DPS) and the Service Center Recognition/Image Processing System (SCRIPS), complementary and integrated into IRS' overall strategy for processing paper documents?

SUMMARY

In summary, Mr. Chairman, IRS' input processing strategy appears to be a high-tech, high-risk, and high-cost venture for which IRS has not yet done the necessary homework to justify committing nearly $3 billion dollars. Further, IRS is planning to spend over $130 million on SCRIPS, another input processing system. Most of the functions of SCRIPS will be duplicated by DPS. To correct these deficiencies, we are recommending that IRS conduct the appropriate comprehensive analyses to justify proceeding with its input processing strategy. We are also recommending that the agency sort out the overlapping functions planned for DPS and SCRIPS to decide the extent to which both systems will be needed.

BACKGROUND

Currently IRS processes about 1.7 billion pieces of paper each year, including about 190 million paper tax returns. Using modified 1950s batch-based computer technology, IRS must process this enormous work load in a timely and accurate manner. In this way, it annually collects and accounts for over a trillion dollars in revenue, works on narrowing the $100-billion tax gap (the estimated amount taxpayers will not pay in income taxes owed), and deals with an accounts receivable inventory that exceeds $100 billion.

The existing processing system is unreliable and unresponsive to the needs of IRS and the taxpaying public. Tax Systems Modernization is IRS' program for making a smooth transition to a new way of doing business using a modernized, electronically-based system for processing tax returns and more rapidly transmitting taxpayer information within IRS, wherever it is needed.

The integrated input processing initiative is a crucial component of the Tax Systems Modernization program. It focuses on the input, storage, and retrieval of tax information and consists of three systems. DPS is considered the heart of IRS' integrated input processing initiative. It is intended to automate IRS' current labor-intensive, paper-based tax return processing system by using state-of-the-art imaging and OCR technology to enter document images and tax data into IRS' systems. IRS plans to use devices to scan paper returns, tax-related documents, and correspondence; these devices will then form electronic images and store them for retrieval and display at workstations located at headquarters and field locations. Data captured by OCR and stored by DPS are to be used by systems, such as those for detecting taxpayers who fail to report income, later in the tax process. In recent testimony before the Senate Committee on Governmental Affairs, we raised concerns about the risks of using this immature technology.[1]

DPS originated in 1988 at which time IRS completed its mission needs statement and began procurement studies. The agency currently plans to award a contract for development of DPS in October 1993, with installation to begin in fiscal year 1996. Through fiscal year 2008, the system is expected to cost about $2.2 billion, or about 80 percent of IRS' estimated costs for the entire input processing initiative.

---

[1]Tax Systems Modernization: Progress Mixed In Addressing Critical Success Factors (GAO/T-IMTEC-92-13, Apr. 2, 1992).

The other two systems comprising this initiative are the cash and electronic management systems. The cash management system is intended to modernize all IRS payment processing, both for paper and electronic remittances. The electronic management system is intended to expand IRS' capability to receive tax data electronically. Together these systems are expected to comprise the remaining approximately 20 percent of the costs for the input processing initiative.

SCRIPS is a related system in that it is planned to have capabilities that will be incorporated in DPS. Although SCRIPS is a modernization project, it is not considered by IRS to be part of the input processing initiative. Its purpose is to replace existing OCR systems at IRS' 10 service centers with upgraded systems. SCRIPS will create images of tax documents and read the numbers on the forms using OCR. It will initially process the same documents the existing equipment processes today. Imaging of documents is not done by the existing systems.

SCRIPS originated in 1987. IRS expects to award a contract for the system's development and installation in November 1992; the system is expected to cost about $132 million. Ultimately, it is to be integrated with, and eventually replaced by DPS.

COST/BENEFIT TRADEOFF ANALYSIS OF INPUT PROCESSING ALTERNATIVES INADEQUATE

With respect to your first question, Mr. Chairman, IRS has not adequately assessed the cost/benefit tradeoffs associated with its input processing initiative. One of the basic tenets of good systems development life-cycle procedures is that agencies embarking on large systems development projects, such as IRS' input processing initiative, must first examine their functional requirements and identify alternatives that could potentially be used for meeting those requirements. This basic tenet is implied in the documentation requirements of Federal Information Processing Standards Publication 64. In addition, under requirements set forth in the Federal Information Resources Management Regulations (FIRMR), agencies must then conduct a comprehensive analysis of these alternatives to identify the alternative that is likely to be most advantageous to the government. According to the FIRMR, the analysis should consider the costs, risk, and magnitude of conversion to the newly developed system. Finally, the FIRMR states that the total estimated cost of each alternative should be calculated using the present value of money. Not using present value techniques tends to understate costs and overstate benefits for most projects. This is because typical systems development projects tend to have most costs earlier and most benefits later in their life cycle.

Since we began our review in January 1992, we made repeated attempts to obtain IRS' cost/benefit tradeoff analyses of input processing alternatives. At various times we were told that such analyses were done, but were either out of date or could not be found. Then, within the past 3 weeks, IRS provided us with several documents, including one showing the agency's analysis of alternative input processing strategies. The analysis compares projected return processing costs at varying levels of electronic versus paper filing.

Our initial review of this document shows that it is totally inadequate for the following reasons:

-- The document is an undated draft consisting mostly of a collection of tables. It lacks a description of its objectives, scope, and methodology, all of which are needed to facilitate its evaluation.
-- The document purportedly covered two analyses, one of which was described, by the vendor who prepared it, as invalid. The second analysis, which did not appear to be included in the document, was described by the vendor as incomplete.
-- Acquisition costs were not discounted using the present value of money.
-- One hundred percent perfect, error free OCR was assumed in the analysis; however, perfect machine character recognition is not currently possible. This assumption excludes the cost of human error correction after character recognition is attempted, which causes the costs for image-based processing to be understated.

Given the complexity and costs of IRS' input processing initiative, a more thorough, strategic analysis of input processing alternatives is needed, and should be regularly updated. From this analysis, IRS should be able to develop a strategically prioritized mix of input processing alternatives and structure its initiative accordingly.

IMPACT OF AGGRESSIVE MARKETING OF ELECTRONIC FILING ON IMAGING COST EFFECTIVENESS UNKNOWN

With respect to your second question, Mr. Chairman, it is impossible to determine, with any precision, what impact aggressive marketing of different kinds of electronic filing would have on the cost effectiveness of imaging. This would require the comprehensive analysis of alternatives that IRS should have conducted in the early phase of its input processing initiative. The long-term analysis we have in mind would take into consideration both markets and technology to tie together:

-- The different types of taxpayers and how they file returns;
-- Which current and potential filing media are likely to be most suitable for each type of taxpayer;

-- What volumes would be involved for each filing medium;
-- How volumes would change for each filing medium over time due to changes in technology, demographics, and marketing strategies including incentives;
-- The costs and benefits of providing these alternative media and the mix of alternatives that would be most advantageous to the government; and,
-- A timetable for phase-in and phase-out of specific alternatives.

To date we have not seen an IRS analysis tying these elements together.

We believe, however, that aggressively marketing different types of electronic input could significantly affect the cost effectiveness of image based processing. Technological advances now make it possible for IRS to receive and enter taxpayer data in a variety of ways. A key advance over recent years is electronic data interchange. This advance allows data to be transferred from taxpayers electronically, directly into IRS' computers. IRS' electronic filing and its current experiment with telefiling are examples of this technology. Paper and its handling costs can be significantly reduced and in some cases, eliminated. In this way paper filing and electronic data interchange are mutually exclusive methods of providing data to IRS. An increase in the volume of electronically filed returns results in a corresponding decrease in the volume of paper filed returns. The costs and benefits of computer systems designed for paper filing and electronic data interchange are directly affected by the volumes of returns these systems process. IRS' statement in the Design Master Plan that electronic transactions are "the most efficient way for the IRS to receive data..." underscores the attractiveness of electronic data interchange.

IRS analysis is needed to accurately determine the point at which the volume of electronic input will reduce paper input such that extensive imaging and character recognition would not be cost effective. If this breakeven point is realistic based on marketing studies, IRS might need to raise the priority of expanded electronic filing to maximize the cost effectiveness of input processing. Under current priorities, expanded electronic filing is planned to be implemented in 1999, three years after DPS. In addition, $271 million is planned for the Electronic Data Interchange/Electronic Management System project life cycle, compared to nearly $2.2 billion for DPS.

IRS does have an existing electronic filing system; however, it has limited growth potential. This system began in 1986 and the agency expects to receive 11 million returns filed electronically through the system this year. This existing electronic filing system, however, is based on older technology and no additional acquisitions are planned after fiscal year 1993. In addition,

there are barriers to electronic filing for taxpayers which IRS could address, for example, by making electronic filing more accessible to the public and devising incentives for all taxpayers and information return filers to file electronically.

CURRENT INPUT PROCESSING STRATEGY RISKY

With respect to your question, Mr. Chairman, about the technical feasibility and realistic nature of IRS' current input processing strategy, we believe there is high risk involved. The risk is due to questions about the feasibility of OCR technology and the realism of implementing such a costly, complex system on such a large scale. Character recognition of handwritten documents is not a mature technology and research is continuing. It cannot yet perform at the speeds and low error rates that IRS requires. Implementation of a technique to enhance this technology's accuracy, called forms redesign and standardization, has not yet begun. In addition, the issue of organizational change due to new business processes, brings into question the realism and escalates the risk of implementing DPS as currently planned. Resolution of these issues is critical to the strategy's success.

The OCR systems that IRS will use must be able to quickly and accurately read handwritten and machine printed characters on a variety of forms. However, this technology is just now being pilot-tested or placed into production at several organizations throughout the nation. For example, the State of Wyoming is using a system to scan and optically read about 100,000 single-sided tax documents this fiscal year. Because no OCR system can read all written and printed characters with perfect accuracy, Wyoming staff must manually correct data read from nearly every form. American Express is testing OCR technology to read the handwritten charges on credit card slips. Although research is continuing, it is difficult to determine when handwritten OCR will become commonplace in large systems. Since IRS currently receives about 100 million handwritten tax returns each year, the risk of using OCR before it is proven is significant. To date, OCR has not been attempted by anyone on IRS' scale.

To maximize the speed and accuracy of OCR technology, different tax forms need to be easily recognized by the system and data on the forms needs to be clear and readable. Standardization of tax forms, by placing the form number in a consistent place on all forms, for example, speeds processing and locates the data on the form for the computer. In addition, placing guide boxes for data entry on forms can separate handwritten characters, preventing character overlap and increasing OCR accuracy. Currently, form identification and character reading is complicated by IRS' acceptance of different versions of the same form. For example, more than 30 types of form 1040 are currently produced by different vendors. In addition, most of IRS' existing forms are

not designed with the guide boxes that are needed for character separation and enhanced recognition accuracy.

To its credit, IRS began organizing a forms standardization project last January. However, the project's stated goal of redesigning over 500 IRS forms in time for DPS' implementation is a formidable one. IRS will have to devote significant resources to manage the forms redesign effort and complete it on schedule.

Introducing new imaging and OCR technology into the workplace will significantly change business processes. IRS has just begun identifying the comprehensive organizational changes that will take place as a result of DPS. IRS calls this process work systems design. As part of this process, IRS must identify new job skills, and develop performance criteria for these jobs. Massive training of current employees must also occur. In addition, IRS will need to redesign workplace policies and procedures. An official from the State of Wyoming stressed the importance of work system design, indicating that even on their small scale, the impact of imaging and OCR on staff, especially more experienced employees, was significant.

IRS has taken some steps. For instance, it has put together a work system design team composed of IRS employees experienced in all aspects of tax return processing. IRS plans to hire a contractor to train and assist the team. Work system design was to begin March 1; however, contract award has been delayed due to procurement problems. For example, review and approval of the vendor was needed, since the vendor chosen had never held a federal contract. IRS expects to resolve these problems in the next couple of months. Work system design must be complete in time for initial DPS implementation in fiscal year 1996.

DPS' AND SCRIPS' PLANNED FUNCTIONS OVERLAP

In answer to your fourth question, Mr. Chairman, the functions IRS plans to perform using SCRIPS would overlap some of those planned for DPS. Both systems are planned to image and perform OCR on some of the same documents, and are to be operational over the same time frames. This is largely the result of slippage in SCRIPS' implementation schedule.

Initially, SCRIPS was supposed to have been implemented by August 1991 and was to have replaced IRS' existing OCR equipment. The documents targeted for conversion to SCRIPS were the same as those currently processed using non-imaging scanners at the IRS service centers. These documents include federal tax deposit coupons, forms 941 and 1040EZ, and information returns. According to IRS, this equipment was to remain in use until 1996-1997, at which time it was to have been replaced by DPS equipment.

According to IRS, SCRIPS now will not be completely implemented until around May 1994. Among the reasons given for this slippage were procurement delays. For example, although the Request for Proposals was forwarded to IRS' contracting office for review in May 1989, that office's initial review did not begin until January 1990.

Thus, SCRIPS' functionality will overlap some of DPS'. In this connection, DPS' imaging and, where applicable, optical data capturing capability is to be applied to all paper documents received by IRS including those forms that are to be processed under SCRIPS. This capability is scheduled to be operational starting in fiscal year 1996, about 2 years after SCRIPS' scheduled implementation. Should IRS proceed with both systems as planned, this overlap would waste taxpayers money and be inconsistent with the good management principles embodied in the Paperwork Reduction Act of 1980. In this connection, the act requires that agency information systems do not overlap each other, or duplicate those of other agencies.

In discussing this overlap with IRS project officials, we were told that IRS was considering modifying the conversion plan for DPS to allow some forms to remain on SCRIPS. They explained that with the slippage in SCRIPS' schedule and its potentially shorter useful life, the benefits to be gained from SCRIPS would not be sufficient to justify the project's costs. Allowing these forms to remain on SCRIPS would increase that project's benefits. It should be noted, however, that this increase would be offset by a commensurate decrease in DPS' benefits, and the wasteful overlap between these two systems would remain.

RECOMMENDATIONS TO THE COMMISSIONER OF INTERNAL REVENUE

In light of the high risk and high cost of IRS' input processing strategy, it is imperative that the Commissioner of Internal Revenue take the following steps to ensure that this strategy is properly justified before committing to its implementation.

-- Conduct a comprehensive analysis to determine the costs and benefits of alternative input processing strategies. This analysis should, at a minimum, take into consideration anticipated technological advances; identify the different types of taxpayers and how they file returns; assess the current and projected filing media that would be suitable for each type; and determine the potential impact of various electronic filing incentives on the requirements for imaging and OCR. This analysis should be updated periodically.

-- Based on this analysis, structure IRS' input processing strategy around the alternative, or mix of alternatives, that are determined to be most advantageous to the government.

-- Reexamine the functional requirements for DPS and SCRIPS and determine whether both systems will be needed and, if so, which system will be used for what so as to eliminate any potential overlap.

- - - - -

This concludes my statement, Mr. Chairman. We will be happy to respond to any questions you or other members of the Subcommittee may have at this time.

(510762)

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20877

Orders may also be placed by calling (202) 275-6241.